CUCS **Resources**

You can subscribe to the CUCS news mailing list at
http://lists.cs.columbia.edu/mailman/listinfo/cucs-news/

CUCS colloquium information is available at
http://www.cs.columbia.edu/lectures

Visit the CUCS alumni portal at **http://alum.cs.columbia.edu**

Columbia University in the City of New York
Department of Computer Science
1214 Amsterdam Avenue
Mailcode: 0401
New York, NY 10027-7003

ADDRESS SERVICE REQUESTED

NON-PROFIT ORG.
U.S. POSTAGE
PAID
NEW YORK, NY
PERMIT 4332

CS@CU

NEWSLETTER OF THE DEPARTMENT OF COMPUTER SCIENCE AT COLUMBIA UNIVERSITY VOL.2 NO.2 SPRING 2006

Members of the Computer Graphics group have been working on interactive methods to produce realistic computer-generated images. The goal is to create images indistinguishable from a real photograph. The image shown here can be rendered at interactive rates, allowing a lighting designer to interactively change both the lighting and view. Note a number of important shading effects, such as soft shadows on the floor, glossy reflections from the bowls, and the realistic appearance of oranges using measured reflectance. The technique is based on a new mathematical framework of triple product wavelet integrals developed by Professor Ravi Ramamoorthi and collaborators. See the accompanying story on Pages 2-3.

Message from the Chair

Henning Schulzrinne, Professor & Chair

Welcome to the spring 2006 edition of the Columbia Computer Science newsletter!

Our Department has been focusing on a number of core areas of strength; we have introduced some of these in past editions of the newsletter, and this edition describes some others. Our graphics group (pgs. 2-3) has grown in depth and breadth over the past few years, adding both distinguished senior faculty such as Prof. Belhumeur and promising junior faculty such as Prof. Grinspun and Prof. Ramamoorthi. We have always prided ourself on the theory group (pgs. 4-5), which now has connections into many other areas of research in the Department, from machine learning to computer security.

Unfortunately, one of the unpleasant realities of academia is that the ability to distribute work via the Internet has made it much easier for students to "outsource" their homework assignments. Prof. Ross recently acted as our local CSI sleuth (pg. 8) and unearthed a plot that will unfortunately likely be repeated elsewhere. It seems increasingly common that I get email from students asking for help with a "networking question". Somehow, they never write back when I ask them which course they are taking.

Each fall, the Department has been organizing a distinguished lecture series, drawing speakers from across the spectrum of computer science (pg. 6-7). As computer science research specializes into ever more sub-disciplines, such talks provide a welcome opportunity to see the best of what is going on in other parts of the field. Our lectures are public, and we invite alumni and friends to join us for the Fall 2006 series. You can subscribe to our colloquium announcement list by visiting our web page *http://www.cs.columbia.edu/lectures*.

The Department has recently upgraded its rather dated phone infrastructure. The article (pg. 10-11) illustrates the necessary planning and hard work, but also that there are opportunities for improving technology that isn't considered all that glamorous.

Like in most CS departments, our undergraduate enrollments have declined from their early-2000 peaks. This decline in computer science enrollments is a widespread phenomenon across the United States; even CNN has recently reported on it.

(continued on page 15)

The Computer **Graphics Lab**

Assistant Professors Ravi Ramamoorthi and Eitan Grinspun head the Computer Graphics Lab, and are involved in a wide range of activities in computer graphics.

In participating media like mist, fog or haze, one sees a number of visual effects like the glows around light sources, and the loss in saturation and contrast on the shading of objects. This complex image can be rendered at 20 frames per second on current graphics hardware and showcases the types of atmospheric scattering effects we can simulate for interactive applications. The technique is based on a novel analytic model for single scattering.

Much of this work is done in collaboration with other members of Columbia's Vision and Graphics Center, such as Prof. Shree Nayar, Prof. Peter Belhumeur, and Prof. Steve Feiner. One of the main themes of our work has been to develop the mathematical foundations of the field, leading to more robust computational algorithms. Here, we provide examples of some of the ongoing projects.

Interactive Photorealistic Rendering

Much of Prof. Ramamoorthi's research focuses on creating realistic computer-generated images, known as rendering. A long-standing goal in computer graphics has been to make images photorealistic or indistinguishable from a real photograph. We seek to capture the effects of natural lighting and shading patterns, such as the soft shadows from the leaves of a tree in skylight, the glints of sunlight in ocean waves, or the shiny reflections from a velvet cushion. In computer graphics, it is important to be able to accurately reproduce these appearance effects, to create realistic images for applications like video games, vehicle and flight simulators, or architectural design of interior spaces. While a considerable amount of progress has been made in rendering realistic images offline, as evidenced by the number of entirely computer-generated movies now released every year, it is still very difficult to accurately model complex illumination and reflection effects in interactive applications like video games. For instance, it is rare to find images in real-time applications with natural illumination from wide-area light sources such as skylight, realistic material properties like velvet, satin, paints, or wood, and shading effects like soft shadows.

Over the past 5 years, in collaboration with a number of other researchers at Columbia and outside, we have developed a number of techniques for high quality real-time rendering, including effects such as soft shadows, complex lighting and material properties, intricate reflections and shading patterns, and the appearance of scenes in scattering media like mist and fog. Images that took hours to render can now be created at several frames per second on modern computers.

Mathematical and Computational Foundations of Visual Appearance

Developing new practical rendering algorithms requires a deeper understanding of the mathematical foundations of visual appearance and lighting simulation. Professor Ramamoorthi and collaborators have made some of the most significant recent contributions in this area, including a novel signal-processing framework for analyzing and representing light reflection, a theoretical framework of generalized triple product integrals for taking complex lighting, glossy reflection and shadowing effects into account for real-time rendering, and new analytic results for modeling the effects of light scattering in participating media like mist, fog, or underwater scenes.

Mathematical and Computational Foundations of Physical Modeling

The behavior of physical systems is typically described by a set of continuous equations using tools such as geometric mechanics and differential geometry to analyze and capture their properties. For purposes of computation one must derive discrete (in space and time) representations of the underlying equations. Researchers in a variety of areas are converging on theories, which are discrete from the start, and have key geometric properties built into their discrete description can often more readily yield robust numerical simulations which are true to the underlying continuous systems: they exactly preserve invariants of the continuous systems in the discrete computational realm. As part of these efforts, Professor Grinspun is developing physical models founded on the core ideas of discrete differential geometry and discrete geometric mechanics. In 2005 he organized the first SIGGRAPH course on Discrete Differential Geometry.

Multiresolution Methods for Geometric Modeling and Physical Simulation

Physical phenomena such as the crushing of a car or the evolution of a storm system are governed by effects ranging from very small to very large scales. Accurately predicting these by resolving the finest scales in a computer simulation is prohibitively expensive. Professor Grinspun and his colleagues study how fine scale information impacts coarse scale behavior and vice versa. In effect "summarizing" these relationships allows the researchers to model coarse scale effects accurately and efficiently without the need to explicitly resolve the finest scales in a computation. A key to this study lies in the careful transfer of structures present in the mathematical models of these phenomena (which in essence have infinite resolution) to the computational realm with its finite resolution and finite computational resources. The methods being developed will allow rapid assessment of overall effects with the ability "to drill down" computationally where additional detail is required.

The **Theory** Group

The theory group includes faculty members Zvi Galil, Jonathan Gross,Tal Malkin, Rocco Servedio, Clifford Stein, Joseph Traub, Henryk Wozniakowski, Mihalis Yannakakis, and other researchers, faculty, and visitors.

Columbia theory group members work in a wide range of different areas. These include the analysis and design of algorithms, coding theory, combinatorial mathematics and graph theory, combinatorial optimization, computational complexity, cryptography, distributed computing, information-based complexity, learning theory, and quantum computation. Information about current visitors, seminars, and events can be found at *http://www.cs.columbia.edu/theory*.

Zvi Galil is the Julian Clarence Levi Professor of Mathematical Methods and Computer Science at the Computer Science Department. He currently serves as the Morris A. and Alma Schapiro Professor of Engineering and Dean of The Fu Foundation School of Engineering and Applied Science. Born in Israel, he was educated at Tel Aviv University and received his doctorate from Cornell University.

His main research interests are in the design and analysis of algorithms, computational complexity and cryptography. He has written over 200 scientific papers, edited 5 books, and has given more than 150 lectures in 20 countries. Dean Galil has served as editor in chief of two journals and as the chief computer science adviser in the United States to the Oxford University Press. He is a Fellow of the ACM and of the American Academy of Arts and Sciences and a member of the National Academy of Engineering.

Jonathan Gross is a Professor of Computer Science. He came to Columbia in 1969 from the faculty of the Mathematics Department at Princeton University and was a founding member of the Computer Science Department in 1979.

His voltage graph construction, a combinatorial analogue of a Riemannian branched covering, has been used by hundreds of researchers in topological graph theory for placing graphs on surfaces. He also introduced the algebraic theme of counting the number of ways that a graph can be placed on a surface. His algorithmic results include the development of a linear-time planarity test for a 2-dimensional polyhedral complex and of the first known polynomial-time algorithm for the maximum genus of a graph.

Professor Gross has also worked in 3-dimensional topology and on the development of sociometrics and experimental design for cross-cultural comparisons in social anthropology. In addition to many papers in mathematics, computer science, and social anthropology, he has also written 14 books, including "Topological Graph Theory" and "Measuring Culture" and has two more books in preparation.

His research awards include an IBM Postdoctoral Fellowship in Mathematical Sciences and an Alfred P. Sloan Fellowship. He was awarded the Outstanding Instructor Award for SEAS in 1991-92 and again in 1993-94. He received the Great Teacher Award for Columbia University in 1994.

Tal Malkin is an Assistant Professor of Computer Science. She received her Ph.D. in Computer Science from the Massachusetts Institute of Technology in 2000, and joined Columbia in 2003 after three years as a research scientist in the Secure Systems Research Department at AT&T Shannon Laboratory. Her research interests are in cryptography, complexity theory, security, and related areas, focusing on developing sound mathematical foundations for applications requiring security and privacy. She studies the computational complexity of cryptographic tasks towards exploring the limits of what is possible to achieve, develops new cryptographic models, and designs practically useful protocols that can be proven secure within these models.

Specific areas of research include relations among cryptographic primitives, black-box reductions and oracle separations, private information retrieval and secure database access, secure multiparty computation, secure approximations, key evolving digital signatures, and algorithmic tamper proof security. She is the recipient of an IBM faculty partnership award and the CAREER award from the National Science foundation.

Rocco Servedio is an Assistant Professor of Computer Science who joined the department in 2003. Prior to coming to Columbia he did a BA in Mathematics and an MS and PhD in Computer Science at Harvard, and was an NSF Mathematical Sciences Postdoctoral Research Fellow at Harvard. His main research area is computational learning theory, with a special emphasis on the design and analysis of provably correct, computationally efficient algorithms for learning rich classes of functions. He has developed algorithms to learn various types of Boolean formulas and circuits, decision trees, and geometrically defined functions such as intersections of halfspaces. Other research interests include noise-tolerant learning algorithms, "attribute-efficient" algorithms that can learn successfully from very small amounts of data, and accuracy boosting algorithms. A recurring theme in his work is the interplay between computational learning theory and other areas of theoretical computer science such as computational complexity theory, public-key cryptography, and quantum computation. His research is supported by an NSF CAREER award, an Alfred P. Sloan Research Fellowship, and by other grant support from the National Science Foundation.

Clifford Stein is a Professor of Industrial Engineering and Operations Research and of Computer Science. He joined the Columbia faculty in 2001, after 9 years as a Professor at Dartmouth College. His research focuses on the design and analysis of efficient algorithms in a wide variety of application areas, including scheduling, network design, and computational biology. He also has a particular interest in the interaction between theory and practice, by producing algorithm implementations based on cutting-edge theory and by applying theoretically sound approaches to new areas.

He is the winner of several prestigious awards including an NSF Career Award, an Alfred Sloan Research Fellowship and the Karen Wetterhahn Award for Distinguished Creative or Scholarly Achievement. He is also the co-author of the textbook "Introduction to Algorithms," with T. Cormen, C. Leiserson and R. Rivest, and the textbook "Discrete Mathematics for Computer Science" with K. Bogart and S. Drysdale.

Joseph F. Traub is the Edwin Howard Armstrong Professor of Computer Science. He came to Columbia in 1979 as founding chair of the Computer Science Department. From 1971 to 1979 he was Head of the Computer Science Department at Carnegie-Mellon University.

He served as founding Chairman of the Computer Science and Telecommunications Board of the National Academy of Sciences from 1986 to 1992 and has served as founding Editor-in-Chief of the Journal of Complexity since 1985.

He is the author, co-author, or editor of nine books and has published some one hundred and twenty papers. His latest book, "Complexity and Information", co-authored with Professor Arthur G. Werschulz, was published by Cambridge University Press in 1998.

Traub and Professor Henryk Wozniakowski started the field of information-based complexity which is now an active research area involving researchers from many countries. A major focus of his current work is quantum computing.

He has received numerous honors including election to the National Academy of Engineering in 1985, the Emanuel R. Piore Gold Medal from IEEE in 1991, selection by the Accademia Nazionale dei Lincei in Rome to present the 1993 Lezione Lincei, and the 1999 Mayor's Award for Excellence in Science and Technology in New York City. In 2001, he received an honorary Doctorate of Science from the University of Central Florida.

Henryk Wozniakowski is a Professor of Computer Science. Wozniakowski and Professor Traub started the field of information-based complexity, which deals with computational complexity of continuous problems using the real number model of computation. He is also interested in quantum computing for continuous problems, and tractability of multivariate problems.

Mihalis Yannakakis is the Percy K. and Vida L.W. Hudson Professor of Computer Science. Born in Greece, he was educated at the National Technical University of Athens and received his doctorate from Princeton University. He was Director of the Computing Principles Research Department Laboratories at Bell Labs, where he worked for more than for 20 years. He briefly joined Avaya Labs and then spent one year as a Professor of Computer Science at Stanford University before coming to Columbia in January 2004.

He has made major contributions to both algorithms and computational complexity in a variety of areas, including database theory, combinatorial optimization, complexity theory, theory of approximability and verification. He has served as Editor in Chief of the SIAM Journal on Computing and as the program committee chair of numerous prestigious conferences. He is a Fellow of Bell Laboratories and a Fellow of the ACM and has received many awards including the Bell Labs President's Gold Award and the Knuth Prize.

2005 **Distinguished** Lecture Series

The Computer Science Department enjoyed another successful Distinguished Lecture Series in 2005. Organized by Professor Tony Jebara and Professor Steve Feiner, the yearly series brings critically acclaimed scientists and pioneers in academia and industry from across the country to Columbia's campus.

Once at Columbia, these distinguished speakers give talks about their research, field questions from the audience, and meet with faculty and students. The lectures are open to other departments and centers across the university, to our alumni and to various affiliated institutions in the Manhattan area. These talks ensure that our faculty and students see first-hand the latest and greatest research emerging outside the Columbia campus. In the fall of 2005, several lecturers visited us and discussed a broad and exciting range of topics ranging from computer science theory to hardware.

These speakers help make this fall an exciting and inspirational one for computer science at Columbia. We are looking forward to next year as we bring more of the greatest researchers from across the world and hear about their latest success stories in computer science.

Our first lecture was September 12th by **Dr. Craig Nevill-Manning** from Google.

Dr. Nevill-Manning is Engineer Director and founder of Google's first remote engineering center located nearby in mid-town Manhattan. His lecture described how Google is organizing the whole world's information.

Google indexes 10 billion documents and has to distribute computations across many machines using fault-tolerant computations and redundancy to sustain the inevitable and constant failures that emerge as they serve the planet's internet search needs. Dr. Nevill-Manning also described and demoed new services being developed by Google such as Google Maps and the Maps API which enable anyone to build applications that leverage Google's satellite mapping technology. Another interesting technology was Google's smart question-and-answering system which is accessible via cell phone text messaging and lets users anywhere look up almost any interesting factoid.

Professor Wayne Wolf from Princeton visited on October 10th.

Professor Wolf is an award-winning author of books in the areas of FPGA, VLSI and embedded systems. He discussed how computer vision algorithms can be integrated into smart cameras using recent advances in embedded computing and hardware. Such algorithms process video directly using on-board camera hardware, track objects (such as people or cars), and produce high-level scene description information. By having multiple distributed cameras, it becomes possible to cover and image large spacious areas, but new practical issues and challenges emerge. For example, the various cameras must communicate with each other to share information, calibrate and synchronize. Professor Wolf discussed a peer-to-peer architecture to solve such problems without using a central server to manage the distributed cameras.

Bill Gates gave a lecture on October 13th; although it was not officially part of the distinguished lecture series, it was a very exciting day across campus and Gates gave a memorable talk.

Arguably the most famous face in computing, William (Bill) H. Gates is Chairman and Chief Software Architect of Microsoft Corporation which grew from his childhood interest in software and programming. Gates met with CS faculty to hear about their research and the key challenges facing computer science. These ranged from proving bounds on computation time to parallelizing software as chip manufacturing nears the 5 GHz clock rate limit. Gates then gave a lecture to over 1400 attendees where he talked about the growing need for students in computer science and applied mathematics. He also showed off new Microsoft R&D including software for managing media (images and video) as well as interactive smart spaces such as desks that synchronize information across various electronic accessories. Gates' talk was followed by an extensive Q&A session as students asked many questions about privacy, open source, standardization and protocols.

Peter Schroder, Professor at the California Institute of Technology and recipient of the 2003 ACM SIGGRAPH award for Computer Graphics Achievement, visited us on October 17th.

He talked about his latest work in reformulating differential geometry and calculus in a discrete setting. Since graphics computation on real computers require discrete data structures, typically numerical approximation of continuous differential geometry has to be used. Professor Schroder's method, however, can give better algorithms for graphically simulating real systems including fluids where the numerical methods are too coarse. He showed examples where his method simulated fluids and smoke while still giving rise to vortices and other complex phenomena that emerge in real fluids yet have been elusive to simulate using traditional methods. The talk was a great combination of elegant mathematics and stunning graphic animations.

Shafi Goldwasser, Professor at the Massachusetts Institute of Technology visited us on October 26th.

She gave a great theory-oriented talk about the possibility of obfuscation with auxiliary input. Obfuscation is a procedure which takes a computer program and efficiently transforms it into another program that still preserves the functionality of the original program yet is unintelligible and difficult to decipher. Professor Goldwasser showed that it would be unlikely that programs of interest can be easily obfuscated and that it is unlikely that general purpose obfuscators can be fabricated. Professor Goldwasser is a two-time winner of the Gödel prize (one of the top distinctions in computer science) and received the ACM Grace Murray Hopper Award in 1996 for outstanding young computer professional of the year and the RSA Award in Mathematics (1998) for outstanding mathematical contributions to cryptography.

Professor Sebastian Thrun, one of the world's leading experts in robotics and Director of the Artificial Intelligence Lab at Stanford University, visited us on November 14th.

A month before his visit, on October 9th 2005, Professor Thrun and his team made history by winning the DARPA Grand Challenge. This was a government-sponsored multi-team effort to build a self-driven robotic car that can cross over 100 miles of desert road as quickly as possible and without any human intervention. Professor Thrun's car, Stanley, finished the 132 mile DARPA Grand Challenge course first, crossing the finish line in just under 6 hours and 54 minutes. We learned first-hand about the technology behind the Stanley smart-car: various GPS, laser and video sensors as well as the 7 Pentium laptop computers that run a variety of probabilistic algorithms for sensing, inference and control. Professor Thrun's work was a great combination of elegant theory in a high-impact real-world application.

Further details of the lecture series (including streaming videos) are available online at: http://www.cs.columbia.edu/lectures

Academic Dishonesty and the Internet

The Internet provides new opportunities for cheating on computer science class projects.

Professor Kenneth Ross

The following piece by Professor Kenneth Ross appeared as a "Viewpoint" article in the October 2005 issue of ***Communications of the ACM.***

It should come as no surprise to learn that dishonest students sometimes submit work that is not their own, either plagiarizing or developing it collaboratively when independent work is expected. An incident in the Computer Science Department at Columbia University during the spring 2004 semester reveals how far dishonest students will go when the opportunity seems easy and risk-free. By sharing my experience here, I hope to draw attention to the new ways some students are using the Internet to cheat on class projects and to highlight the fact that such cheating can (with effort) be detected.

The programming project I created and assigned to an introductory database systems course involved writing code for bulk-loading a B+ tree. Two weeks into the project, I was surprised to find an anonymous student posting on the class discussion board. The author reported performing a Google search on the phrase B+ tree and to his (and my) surprise turned up a suspicious hit on the fourth screenful of results—a page at www.rentacoder.com entitled "B+ tree bulk loading" with several paragraphs apparently copied verbatim from my project.

Some student was trying to subcontract his project for the class. The posting did not reveal any identifying information about its author. I called www.rentacoder.com on the phone, but the company was not helpful, refusing to reveal anything about posters without a court order. The poster in this case was doing nothing illegal (except perhaps violating the copyright on my project description), so this avenue seemed unproductive.

I told the class what I had learned and posted an announcement to the class discussion board explaining that I was aware of the posting at www.rentacoder.com and considered it an extremely flagrant form of cheating. Sure enough (it seemed to me), the bidding process on www.rentacoder.com was canceled that day. I thought this would be the end of it, that the student had been scared off and would now refrain from cheating.

I was wrong. Toward the end of the course in April, I visited www.rentacoder.com again. While the original call for bids was still canceled, a second call, entitled "bulk loading 2," had been posted by the same person as a private auction, that is, one targeted at a particular coder (probably someone who had responded to the earlier posting). Money had apparently changed hands ($50), and the poster had given positive feedback about the quality of the coding job.

Additionally, the posting now gave a piece of identifying information about the poster, namely that his address was in Miami, FL. However, I wasn't ready to jump to conclusions. Many of our students in New York might originally be from Florida. Also, several students were taking the class remotely via the Web. Moreover, the poster may not have been honest in listing Miami as his address. The listing was part of the www.rentacoder.com trust-establishment process in which a contractor might gain some confidence that a poster is legitimate and the contractor will be paid. My guess was that the Miami location was automatically generated from the address on the credit card the poster used for payment.

At this point, all student projects had been submitted, so I tried to find the problematic one by inspecting the submissions. While I wasn't sure what kind of suspicious code to look for, I did expect I'd know it when I'd see it. Looking through the 82 projects certainly wasn't appealing.

I had, however, fortuitously changed the project specifications early on to make it somewhat easier; a parameter that was supposed to be entered on the command line could now be a compile-time parameter. I reasoned that the student who posted the bid would have either posted the project before the change or not paid attention to the change. Approximately nine submitted projects allowed the runtime parameter, and the teaching assistants forwarded them to me electronically.

It was relatively easy to spot the "bad" project due to several problematic features:

- **Written with an overly elaborate use of C++.** For example, the class "B-tree" was defined, and a subclass "B+-tree" was inherited from it. I hadn't asked anything about B-trees, just about B+ trees;
- **Far more general than it should have been.** I had asked for a bulk-loading algorithm and a search algorithm, but methods were also implemented for insert and delete;
- **Coded wrong.** The algorithm implemented was not, in fact, bulk loading but merely repeated insertions; and
- **Ungrammatical comments.** Even though the student was apparently a fluent English speaker.

It also turned out that the student who had submitted the project was indeed a remote student with a Miami-area telephone number.

I asked the dean of Columbia's Engineering School how to proceed. He told me to call and confront the student with the evidence of cheating which I did. He was reluctant at first to admit any wrongdoing as I described the evidence. He defended his elaborate use of C++ on the basis of having studied C++ in a previous course. He defended the overly general code on the grounds that he had developed a B-tree package for a previous course and had now reused it in my course. He ascribed the wrong implementation of bulk-loading to being behind in the course, including having not yet watched many of the course videos. And he claimed the bad grammar was due to being in a hurry.

However, once I said I had seen a posting on www.rentacoder.com from someone in Miami, he admitted to having made the initial posting. He said he canceled it and didn't use the results but had no rejoinder when I pointed out the second private bid request. At that point he acknowledged using www.rentacoder.com to subcontract the class project. He later sent an email message apologizing for the incident, describing various personal problems and asking for leniency.

I referred the case to the dean for a disciplinary hearing. The student subsequently received a disciplinary penalty and an F for the course. In light of this and other recent cases, the department updated its academic honesty policy (*www.cs.columbia.edu/education/honesty*).

I refrain from sweeping generalizations about the scale of Internet-based cheating on the basis of this one experience. It does, however, seem like students are regularly posting homework on the Net looking for solutions from others. A quick visit to www.rentacoder.com (or other such sites) will verify that the postings I noticed do not represent an isolated incident, although catching an offender is probably rare. We've clearly succeeded in teaching our students about outsourcing.

I've learned that instructors must be aware of the sophistication of dishonest students and match that sophistication in detecting their dishonesty. Cheating will never be eliminated. However, once students hear about even one case in which such a scheme is exposed, that scheme is much less likely to be used again. After all, the penalties for substantive incidents of cheating are likely to be severe.

Perhaps students searching for Internet-based subcontractors will find this column among the hits (it includes most of the right keywords) and appreciate the risk they are taking.

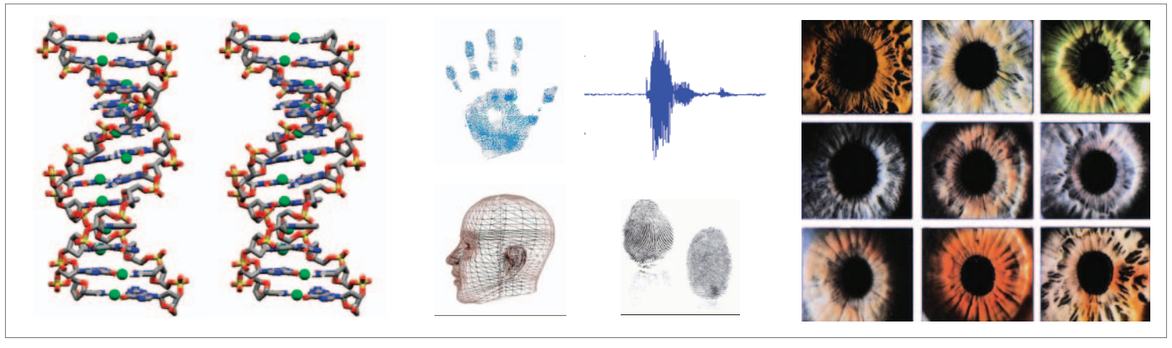
I've deliberately withheld the name of the student, using "he" to refer to this person rather than the awkward "he/she"; the student's gender should not be inferred from this usage.

New Computer Science Courses in 2005-2006

Professor Bellovin taught COMS 6998-7, **Anonymity and Privacy**, in the Spring 2005 semester; it is now a permanent course and is being offered as 6184 in the Spring 2006 semester. Students read and presented a variety of papers and other primary source material. Readings ranged from court opinions and statutes to not-yet-published research papers. In addition, each student researched, wrote up, and presented a topic of his or her choice.

In the Fall 2005 semester, **Professor Bellovin** taught COMS 4995-2, **Introduction to Security**, both in person and via CVN. It is a complement to COMS 4180, "Network Security", and focuses on the architectural and engineering aspects of secure system design.

Professor Peter Belhumeur taught COMS 4995/6998, **Biometrics**, in the Fall 2005 semester. The earliest known use of biometrics dates back to the 7th century during China's Tang Dynasty; during this period fingerprints were used to sign and validate contracts. Over the last century, biometrics—the science for determining a person's identity by measuring his/her physiological characteristics—has grown enormously. Technologies are being developed to verify or identify individuals based on measurements of the face, hand geometry, iris, retina, finger, ear, voice, speech, signature, lip motion, skin reflectance, DNA, and even body odor. This course explores the latest advances in biometrics as well as the machine learning techniques behind them; students learn how these technologies work and how they are sometimes defeated.



CUCS Converts to **VoIP**

The Department of Computer Science has a long history of research and development in networking.

Daisy Nguyen

Dennis Shim

John Petrella

Mark Yeun

Professor **Henning Schulzrinne**

Professor **Yechiam Yemini**

For example, Professor Yemini's work led to the formation of the lead vendor of voice-messaging systems and to the first commercial system for autonomic diagnosis and management of network failures. As a recent example, Professor Schulzrinne's Internet Real-Time Lab has been very active in developing protocols and applications for voice-over-IP (VoIP), in particular the core components for setting up calls and media transport. The most popular standardized call signaling protocol is SIP (Session Initiation Protocol), co-authored by Prof. Schulzrinne. This now-ubiquitous protocol allows voice-enabled devices to communicate over standard IP networks. Given the interest and local expertise, and an aging analog phone system, the Department has been on a path to converting to VoIP. Unlike most other organizational units in the University, the Department had been running its own small Nortel Meridian PBX, with its own set of numbers, for its 100 or so extensions since the early 1990s, while the University was still using Centrex services. The PBX was cost effective, but had reached the limits of its expansion and lacked modern conveniences, such as voice mail or caller ID.

The CS department was an early adopter of voice-over-IP, deploying, in 2001, a SIP software suite originally developed in the IRT Lab. The deployment was initially motivated by running out of capacity on our PBX. Making further investment in the analog system was not a practical solution. To compensate, PBX capacity was extended through VoIP. The SIP-based VoIP software suite known as CINEMA was connected to the Nortel Meridian PBX via a Cisco 2620 SIP-enabled gateway router. CINEMA had been licensed to SIPquest, a Columbia-initiated start-up venture. New users were given Cisco SIP phones, and features such as voicemail and conferencing were also made available.

However, maintaining and upgrading a research platform for department-wide operational use became difficult, particularly as SIPquest shifted to an OEM business model. Working with Emergent Communications, Inc. (Emergent), a VoIP system integrator, Prof. Schulzrinne and the CS department IT (CRF) staff came up with a viable long-term IP-based voice communications solution for the department.

The process started with the issuance of a RFP to several leading VoIP equipment vendors in March 2005. Responses and reviews of various proposals were completed by June and a recommendation was received. Emergent management approached 3Com with a proposal to set up a 3Com system for the CS department because the 3Com VCX IP Telephony Solution and Convergence Application Suite offered the best combination of functionality, reliability, ease of operation and administration. The 3Com system contains software applications that were licensed from SIPquest.

3Com accepted the proposal and donated equipment to the CS department. Engineers from 3Com also donated their time to set up the entire system. The hardware included one VCX VoIP gateway, two SIP proxy servers (for redundancy), one IP conferencing and presence server, a 3Com-Tipping Point intrusion prevention system, two analog media gateway, 20 phones and an attendant console. In addition, 3Com donated the necessary software licenses for a full deployment.

The VCX system can scale far beyond our departmental needs, supporting up to 1,500 users for voicemail, for example. For reliability, the VCX system provides a load sharing, hot standby primary/secondary redundant solution for VCX IP telephony and intelligent mirroring for IP messaging, both with automatic fail-overs and recovery. In addition, the 3Com system can interoperate with third-party equipment. During the transition, the 3Com system worked seamlessly with a Cisco VoIP gateway router and currently uses than one hundred Cisco 7940 and 7960 SIP phones which remained from the previous VoIP system in the CS department.

Since we had to ensure continuous phone service to the Department, the conversion took place in three phases, starting in late August and ending in early November 2005, starting with a testing phase, followed by a hybrid environment, and then cutting over to an all-VoIP environment.

3Com had a solid implementation plan that helped get the department through the migration. 3Com and Emergent teams worked side by side with CRF in every phase. In the final phase, every single one of the 200 or so phones and fax machines in the department was tested in one long night. All the phones worked smoothly, but the fax machines were more complicated; an all-nighter was required to get every analog fax machine to work smoothly via the Sipura VoIP adapter.

After five months, the 3Com system continues to work well. One of the features made possible with the new system is that we now assign phone numbers both to rooms and to individuals, since the IP phones can handle multiple "line presences". When a graduate student or staff member moves to a new office, they can maintain their existing phone number. If they occupy the office by themselves, they just take their phone with them and plug it back into the campus network in their new office.

The new system has other useful features. A popular feature is the ability to connect to the system via soft phone, so that traveling staff and faculty can call any New York City number for free from their laptop. Voice mail is now delivered as an email attachment. Up to six people can share one desk phone, but each user gets his or her own phone number and personal voice mail box. Incoming calls can follow users to other phones. We plan to integrate presence and IM into the system in the future. We may also forego our traditional T1 digital circuit to the phone system and connect directly to a phone gateway provider.

Overall, the transition to an IP-based phone infrastructure allowed us to offer better services to our faculty, staff and graduate students, while maintaining the familiar feel of a hardware-based phone system.

Credits: The 3Com team consisted of Erik Papir, Tom Sammon, Jeff Enter, Bob Briganti, and Roy Bohl; the Emergent Team of Mike Tedder, Santo Pittsman, and Stephen Tarzia. The CRF team (John Petrella, Mark Yeun and Dennis Shim) was led by Daisy Nguyen. The Department thanks 3Com for their generous donation.

Blaine Bell
Advisor: Steven Feiner
View Management for Distributed User Interfaces

Abstract: We present a new approach to designing user interfaces based on view management. View management refers to layout decisions that determine spatial relationships between objects, taking into account visibility constraints that allow applications to manage what users see. Our techniques are used to satisfy these constraints by controlling what is seen in a wide range of user interfaces: from 2D desktops to 3D immersive environments, from view-only presentations to interactive techniques, and from single user to collaborative situations. Screen space is a difficult resource to utilize properly in user interfaces. Poorly designed user interfaces, consisting of overlapping windows, unwanted popups, and unused screen space, occur on many different types of displays, such as PDAs, laptops, cell phones, and wall-sized displays. We avoid these problems by providing techniques for representing and managing unused screen space to avoid overlapping when possible. Applying these techniques to 3D user interfaces imposes visual constraints on the placement of objects, such as labels and annotations, relative to the 2D visible projections of 3D objects on the view plane. We develop techniques to handle issues such as performance and temporal stabilization, and we create guidelines to help ensure that labels and annotations behave well across a wide range of situations. We have applied our techniques to augmented reality, in which information is visualized directly within the context of the real world by overlaying graphics onto what is seen. We develop tools, including an annotated situation-awareness aid based on a world in miniature, that make it easy to visualize information about the environment, including parts of the environment that might not be completely visible from their current location. We extend these ideas further into distributed interactive environments. In a world encompassing a wide range of display technologies, we are interested in how people will access, use, and interact with information and each other. All of this work presented has been developed (or redeveloped) using an efficient rule-based programming technique we call Data Programming, which is designed for standalone and distributed development and has given us the flexibility to investigate a wide range of scenarios in these environments.

Noemie Elhadad
Advisor: Kathleen McKeown
User-Sensitive Text Summarization

Abstract: In this thesis, we present a user-sensitive approach to text summarization. One domain which would highly benefit from tailoring summaries to both individual and class-based user characteristics is the medical domain, where physicians and patients access similar information, each with their own needs and abilities. Our framework is a medical digital library for physicians and patients. We describe a summarizer, which generates summaries of findings in an input set of clinical studies. When a physician is treating a specific patient, he's looking for information relevant to the patient's history and problems. The summarizer takes the user's interests into account and presents only the findings pertaining to a user model, as approximated by an existing patient record. The same synthesis of information can also be of interest to the patient. The summarizer predicts which medical terms used in a text will be too technical for patients, and augments it with appropriate definitions when necessary.

We adopt a generation-like architecture for our summarizer. However, because our input is textual and not semantic, new

challenges arise. We operate over a content representation hybrid between full-semantic and extracted phrases. Our content organization strategy is dynamic and data-driven. This is in contrast to most summarizers which use no explicit strategies to order information extracted from several input documents. The result is more readable, coherent output. To generate the actual summary, the summarizer makes use of aggregation and phrasal generation. The result is a concise and fluent summary.

One key challenge when it comes to adapting a text for a different audience is identifying the bottleneck for reader comprehension. We analyzed corpora of technical and lay medical texts and qualified differences. We identified the presence of difficult vocabulary as the major obstacle to comprehension for lay readers. We designed an unsupervised method to predict which terms are incomprehensible for lay readers and provide the user with appropriate definitions.

Our methods are grounded on corpus analyses and feasibility studies conducted with physicians and consumers of health information. To assess the value of our work, we evaluated our summarizer both intrinsically and extrinsically. Our task-based evaluation conducted with physicians at the ICU demonstrates that personalized summaries help physicians access relevant information better than generic summaries. Evaluation with lay readers shows that our method to augment technical medical texts improves readers' comprehension significantly.

Suhit Gupta
Advisor:
Gail Kaiser
Context-Based Content Extraction of HTML Documents

Abstract: Web pages may often contain "clutter" (defined by us as unnecessary images, navigational menus and extraneous links) around the body of an article that may distract a user from actual content. Extraction of useful and relevant content from web pages has many applications, including speech rendering for the visually disabled, cell phone and PDA browsing, and text summarization. Most existing approaches to making content more directly accessible involve changing font size or removing HTML and data components such as images, which may take away from a webpage's inherent look and feel. Unlike Content Reformatting, which aims to reproduce the entire webpage in a more convenient form, our solution directly addresses Content Extraction. We introduce Crunch, a framework that employs an easily extensible set of techniques, for enabling and integrating heuristics concerned with content extraction from HTML web pages. Crunch is implemented as a transparent web proxy and is practically usable by end-users. We use DOM tree based content extraction rather than directly processing HTML as flat files. Crunch is a versatile solution, allowing programmers and administrators to add heuristics to the framework. These heuristics act as filters that can be parameterized and toggled to perform the content extraction. Crunch reduces human involvement in the application of thresholds for the heuristics by automatically detecting and utilizing the content genre (context) of a given website. Genre detection is accomplished via the use of frequency distributions of words associated with the webpages. These distributions are used to improve the extraction process by comparing them to previously known results that work well for certain genres of sites and utilizing those settings. We have measured the usability and performance of the content extraction proxy in terms of the quality of the output generated by the heuristics that act as filters after the proxy has inferred the context of a webpage. Ultimately, we show that rather than going with current approaches that are prepackaged "one size fits all" and programmer controlled, going with a more adaptable approach will produce a more content-full result.

Marek Kwas
Advisor: Henryk Wozniakowski
Quantum Algorithms and Complexity for Certain Continuous and Related Discrete Problems

Abstract: We present an analysis of two computational problems. The first problem is discrete quantum Boolean summation. This problem is a building block of quantum algorithms for many continuous problems, such as integration, approximation, differential equations and path integration. The second problem is continuous multivariate Feynman-Kac path integration, which is a special case of path integration.

The quantum Boolean summation problem can be solved by the quantum summation (QS) algorithm of Brassard, Hoyer, Mosca and Tapp, which approximates the arithmetic mean of a Boolean function. We improve the error bound of Brassard et al. for the worst-probabilistic setting. Our error bound is sharp. We also present new sharp error bounds in the average-probabilistic and worst-average settings. Our average-probabilistic error bounds prove the optimality of the QS algorithm for a certain choice of its parameters. The study of the worst-average error shows that the QS algorithm is not optimal in this setting; we need to use a certain number of repetitions to regain its optimality.

The multivariate Feynman-Kac path integration problem for smooth multivariate functions suffers from the provable curse of dimensionality in the worst-case deterministic setting, i.e., the minimal number of function evaluations needed to compute an approximation depends exponentially on the number of variables. We show that in both the randomized and quantum settings the curse of dimensionality is vanquished, i.e., the minimal number of function evaluations and/or quantum queries required to compute an approximation depends only polynomially on the reciprocal of the desired accuracy and has a bound independent of the number of variables. The exponents of these polynomials are 2 in the randomized setting and 1 in the quantum setting. These exponents can be lowered at the expense of the dependence on the number of variables. Hence, the quantum setting yields exponential speedup over the worst-case deterministic setting, and quadratic speedup over the randomized setting.

Ani Nenkova
Advisor: Kathleen McKeown
Evaluation of Top-k Queries over Structured and Semi-structured Data

Abstract: Recent years have seen unprecedented interest in news aggregation and browsing, with dedicated corporate and research websites becoming increasingly popular. Generic multi-document summarization can enhance the users' experience with such sites, and thus the development and evaluation of automatic summarization systems has become not only research, but a very practical challange. In this thesis, we describe a general modular automatic summarizer that achieves state of the art performance, present our experiments with rewrite of generic noun phrases and of reference to people, and demonstrate how distinctions such as familiarity and salience of entities mentioned in the input can the automatically determined. We also propose an intrinsic evaluation method for summarization that incorporates the use of multiple models and allows a better study of human agreement in content selection. Our investigations and experiments have helped us to understand better the process of summarization and to formulate tasks that would lead to future improvements in automatic summarization.

David Olshefski
Advisor:
Jason Nieh
Measuring and Managing the Remote Client Perceived Response Time for Web Transactions using Server-side Techniques

Abstract: As businesses continue to grow their dependence on the World Wide Web, it becomes increasingly vital for them to have accurate, quantitative measures of the performance of their web services. Specifically, the focus of web server performance has shifted from throughput and utilization benchmarks to guaranteeing delay bounds for different classes of clients. Providers of web services are faced with the challenge of providing differentiated services that guarantee bounds on client perceived response times while at the same time maximizing throughput and minimizing cost. In order for a web site to guarantee delay bounds for its clients, it must be able to determine the response time, *as perceived by the remote clients*—in real time, with provable accuracy. This information can then be used as one of the key inputs to a management system that can verify compliance with service-level objectives and tune system parameters for performance. Unfortunately, the problem of obtaining an accurate measure of client perceived response time has remained a key factor in preventing delay bounded web services from being realized.

This thesis shows that it is possible to determine the remote client perceived response time for web transactions using only server-side techniques and that doing so is useful for the management of latency based service level agreements. We present several novel mechanisms to achieve client perceived response times using server-side techniques and show how they can be used in system management to obtain service level agreements. This thesis addresses several of the key problems from an engineering, modeling and theoretical perspective. First, we present a novel modelling algorithm named *Certes*, that can estimate the connection establishment latencies *as perceived by the remote clients*. Certes executes within the operating system of the web server and is shown to accurately estimate the response times experienced by the remote client in the presence of admissions control drops. The second novel contribution of this work entails the design and development of an intelligent traffic monitor named *ksniffer* which is capable of determining the page view response times experienced by a remote client. Implemented as an appliance that is placed in front of the web server it uses novel algorithms for inferring the remote client perceived response time *on a per pageview basis*. The third novel contribution of this work, *Remote Latency-based Management*, entails extensions to ksniffer which manipulate the latencies experienced by the remote client by manipulating the packet traffic into and out of the web server complex. Our system tracks the progress of each page download in real-time, as each embedded object is requested, allowing us to make fine grained decisions on the processing of each request as it pertains to the overall page view latency.

Krysta Svore
Advisor: Al Aho
Software Tools and Failure Thresholds for Reliable, Scalable, Fault-Tolerant Quantum Computation

Abstract: Quantum computation has the potential to solve certain mathematical and physical problems faster than a classical computer. However, one of the key challenges to realizing a quantum computing device is the control of systematic errors and maintaining the coherence of the quantum state. In addition, once a quantum computer is built, we must be able to communicate a quantum algorithm to the physical device in a fault-tolerant manner. In this dissertation, we take several significant steps towards the realization of a quantum computer. We develop a framework for use with any proposed technology to map a quantum algorithm into fault-tolerant machine instructions. Within this framework, we combat errors and decoherence by introducing quantum error correction and fault tolerance into a quantum circuit. We analyze specific fault-tolerant circuitry on two general architectures, nonlocal and nearest-neighbor, to determine failure thresholds of the physical circuit components. These failure thresholds allow physicists to evaluate technology proposals. Our goal is to provide design tools, failure thresholds, and fault-tolerant constructions to combat the faultiness of quantum computation.

Lijun Tang
Advisor:
John Kender
Semantic Content Analysis and User Interface for Instructional Video Indexing

Abstract: Video semantic content (high-level concepts) can be presented through many different visual presentations according to different video genres. The process of extracting the semantic content is complex, because it requires domain knowledge or user interaction. We concentrate our research on semantic content analysis for instructional videos, where the video text (including the instructor's handwriting) provides significant amounts of high-level information.

We first extract text frames from instructional videos, remove redundant text frames, then enhance the visual quality of these text frames through super-resolution from multiple adjacent frames. These text frames provide image-level summaries of instructional videos. We further extract high-level concepts by topic word spotting techniques: We construct a topic word vocabulary from supplement materials like course syllabus, table of contents (TOC) of the text book, electronic slides, etc. Then starting from a dynamic programming analysis by synthesis technique, we implement both structure analysis based and statistical handwriting recognition algorithms and explore the optimal handwritten word recognizer for videotext recognition. Using the optimal handwritten word recognizer, we match the topic words in the vocabulary against the handwritten word images segmented from text frames. We then construct a topic-frame matrix, and by querying this matrix with topic words in TOC, we are able to determine which chapter/topic is most likely discussed in each section of video. We designed five novel graphic user interfaces (UI) to present the results and conduct user studies to evaluate the proposed UI and approaches.

Department News & Awards

Matei Ciocarlie, **Matthew Burnside**, **Chris Murphy, Agustin Gravano** and **Marcio Buss** were named as "extraordinary TAs" for the Spring 2005 semester, based on the evaluation of students in their classes.

Professor **Peter Allen** and **Dennis Fowler**, M.D. (Surgery) have received a two-year, $425K NIH Exploratory/ Developmental Research Grant for Insertable Imaging and Effector Platforms for Surgery. The grant is to construct small, mobile, multi-function platforms that can be placed inside a body cavity to perform robotic minimal access surgery. The robot will be based upon an existing prototype device developed at the Columbia Robotics Lab.

Ph.D. student **Matei Ciocarlie** was chosen as the second place winner in the CanestaVision 3D Vision Contest. Matei is a PhD student in Professor Peter Allen's Robotics lab. The prize includes a $5,000 cash award and an electronic perception development kit worth $7,000. Matei's entry was a real-time "Eye-in-Hand" range sensor for robotic grasping. Matei was one of ten finalists, who were then given six months to develop their 3D vision application.

Rean Griffith, a PhD student working under the supervision of Prof. Gail Kaiser, won the distinguished IBM PhD fellowship, a highly competitive award. Rean is a 7th semester PhD student who had previously received his MS at Columbia and his BS from the University of the West Indies, Barbados, and has worked as an intern the past two summers at IBM Almaden as well as at Microsoft. He has to date published or had accepted for publication half a dozen papers joint with IBM Watson researchers, including an IEEE Transactions journal article and a book chapter in a forthcoming CRC Autonomic Computing 'handbook'. His tentative thesis proposal title is "An Approach to Retro-fitting and Evaluating the Self-Healing Capabilities of Legacy Systems". Rean's thesis topic concerns developing technologies to dynamically inject self-healing capabilities into legacy software systems without available source code, to perform adaptations while those systems continue running, and devising benchmarks to qualitatively and quantitatively compare alternative autonomic self-healing algorithms that can be injected in in this fashion.

The Second Edition of the textbook "Graph Theory and Its Applications" by Professor **Jonathan Gross** was published in the fall.

Professor **Angelos Keromytis** was awarded a two-year grant under the NSF CPA program titled "Retrofitting A Flow-oriented Paradigm in Commodity Operating Systems for High-Performance Computing".

Undergraduate students **Claire Lackner** and **Catherine Lennon** were selected for Honorable Mention for the Computing Research Association's Outstanding Undergraduate Award 2006, sponsored by Mitsubishi Electric Research Labs. They are part of a select group of only 20 female undergraduates in the country that were honored with the Outstanding Undergraduate Award this year.

Professor **Tal Malkin** and PhD students **Homin Lee** and **Andrew Wan** were featured in a December www.nerve.com column on "Sex Advice From Cryptographers." Nerve is a "smart, honest magazine on sex, with cliche-shattering prose and fiction as well as striking photographs of naked people that capture more than their flesh." Each installment of the "Sex Advice From..." column features sex advice from members of a particular niche of society; previously featured groups include "Jiu-Jitsu Fighters," "New Parents," "Equestrians" and "Bike Jousters."

Angelos Keromytis and **Vishal Misra** were each promoted to the rank of Associate Professor. **Kenneth Ross** was promoted to the rank of Professor.

Professor **Shree Nayar** has received many highly competitive grants recently. These include "Vision through Rain and Snow" from the National Science Foundation; "A Physical Approach to Underwater Imaging" from the Office of Naval Research; "Advanced Imaging Technology" from Sony Research Laboratory, and "Methods for Controllable and Flexible Imaging" from the Office of Naval Research. Together with Professor **Ravi Ramamoorthi**, Shree also received a grant for "Fast and Accurate Volumetric Rendering of Scattering Phenomena in Computer Graphics" from the National Science Foundation.

Ph.D. student **Shaya Potter** and Professor **Jason Nieh** received the Best Student Paper Award at the 19th Large Installation System Administration Conference (LISA 2005) held last week in San Diego, CA for their paper titled: "Reducing Downtime Due to System Maintenance and Upgrades".

Professor **Henning Schulzrinne** was elected as an IEEE Fellow for his contributions to the design of protocols, applications, and algorithms for Internet multimedia.

The textbook "Discrete Mathematics for Computer Science" by **Kenneth Bogart**, **Professor Cliff Stein** and **Scot Drysdale** was published this summer.

Professor **Cliff Stein** was the Program Committee Chair for the Symposium on Discrete Algorithms, held in Miami in January, 2006.

Dr. **Owen Rambow**, Senior Research Scientist in Center for Computational Learning Systems, has been elected to a two-year term as chair of the board of the North American Chapter of the Association for Computational Linguistics.

Professor **Sal Stolfo**'s startup CounterStorm was featured in NetworkWorld. CounterStorm unveiled CounterStorm-1, a security appliance that protects network resources early on in an attack by shutting down network access.

Professor **Mihalis Yannakakis** was awarded the Donald E. Knuth Prize at the 46th Symposium on Foundations of Computer Science (FOCS 2006). The Knuth prize for outstanding contributions to the foundations of computer science is awarded every 1.5 years by the ACM Special Interest Group on Algorithms and Computing Theory (SIGACT) and the IEEE Technical Committee on the Mathematical Foundations of Computing (TCMFC). The Prize includes a $5000 award and a $1000 travel stipend (for travel to the award ceremony) paid by ACM SIGACT and IEEE TCMFC. The Prize is awarded for major research accomplishments and a sustained record of high-impact, seminal contributions to the foundations of computer science over an extended period of time.

Vladimir Vapnik, who is a member of Columbia's Center for Computational Learning System and is also affiliated with the Computer Science department, has been elected to the National Academy of Engineering "for insights into fundamental complexities of learning and for practical and widely applied machine learning algorithms."

The third Theory of Cryptography Conference (TCC 2006) was held at Columbia in early March. TCC deals with the paradigms, approaches and techniques used to conceptualize, define and provide solutions to natural cryptographic problems.

The ninth International Conference on Practice and Theory in Public Key Cryptography (PKC 2006) was held at Columbia in late April. Professor **Tal Malkin** is the General Chair and Senior Research Scientist **Moti Yung** is the Conference and Program Committee Chair. PKC has been the main IACR annual workshop focusing on all aspects of public-key cryptography, and has attracted papers from world-renowned scientists in the area.

Alumni News

Regina Barzilay (Ph.D. '02) was selected by IEEE Intelligent Systems as one of the "IEEE IS Ten to Watch". She will be featured in a forthcoming special issue published by IEEE Intelligent Systems published to coincide with the 50th Anniversary of the Dartmouth Workshop (generally considered the birthplace of modern AI). Prof. Barzilay, who received her PhD in 2002 under the supervision of Prof. Kathy McKeown, is currently an assistant professor at the Computer Science and Artificial Intelligence Laboratory at MIT.

Blaine Bell (Ph.D. '05) finished his dissertation under Prof. Steven Feiner in 2005. He is now working at Schrodinger, a company started by Columbia chemistry professor Richard Friesner that develops software for the pharmaceutical industry.

Noemie Elhadad (Ph.D. '05) finished her dissertation under Prof. Kathy McKeown in 2005. She is now an Assistant Professor in the Department of Computer Science at City College, CUNY.

Gaurav Kc (Ph.D. '05) finished his dissertation under Prof. Alfred Aho in 2005. He is working at Google in Mountain View.

Simon Lok (Ph.D. '05) received the "Emerging Talent" award at the Urban Visionaries Fourth Annual Award Dinner sponsored by Cooper Union (see http://www.cooper.edu/urbanvisionaries/) on March 28. Simon is Chief Scientist and Chairman of the Board of Directors of Lok Technology, Inc., an innovator in secure, integrated network infrastructure appliances for wireless and wired networks. The dinner honors "distinguished figures whose outstanding contributions to city life exemplify the values championed by The Cooper Union."

Ani Nenkova (Ph.D. '05) finished her dissertation under Prof. Kathy McKeown in 2005. She is now a post-doctoral research associate in the Department of Linguistics at Stanford University, working with Dan Jurafsky.

Vlad Shchogolev (CC '04) writes: "I'm enrolled part time in the graduate program at Columbia. I'm also working with the Machine Learning lab. Professor Jebara and I recently submitted a paper to ICML. The rest of the time I work as a software engineer at Google. None of my work is publically visible yet, but that may soon change."

(Message from the Chair
continued from page 1)

Our graduates seem to have no difficulty finding good jobs, both in the New York City metro area and in other parts of the country, and, from our anecdotal evidence, seem to like their first jobs. As a computer science community, we have a very limited understanding of what drives students and their parents to choose among majors and what accounts for this decline. I suspect that it is a combination of a perceived "cool" factor, news reporting on disciplines and anticipated long-term job stability. Reporting on computer science-related topics seems to focus on negative aspects such as the latest privacy breaches, compared to the potential miracle cures promised by gene splicing. This is reflected in the Columbia Engineering major statistics. Columbia's engineering school applications continue to increase, but students seem to now migrate to biomedical, mechanical, and civil engineering, as the currently largest majors. Having better data, rather than just anecdotes, on what CS graduating students do, how they perceive their career and whether they would recommend their choice to their friends and children might help us to take the initiative, be it in dispelling misconceptions among high-school guidance counselors and students or in changing the CS curriculum. Gathering such data would seem to be worthwhile initiatives for ACM or CRA, just like the Taulbee survey has helped to improve data-based planning at the PhD level. It is clear that computer science as a discipline continues to be attractive. As two local examples, our CS I course (COMS 1004) taught by Prof. Adam Cannon has grown to more than 105 students this year, and our new Matlab course is likely to become a required part of other engineering programs in the School. It has paid off that Prof. Cannon has made the course accessible to our very diverse student population, including students who seem to have written their first program in first grade and others whose computer experience is limited to Word and IM.

I'd welcome your insights and thoughts on how enrollments in Computer Science can once again reflect the excitement and promise of the field. As always, we are eager to hear from our alumni.